PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Potential Role of Particulate Matter in the Spreading of COVID-19 in Northern Italy: First Observational Study based on Initial Epidemic Diffusion
AUTHORS	Setti, Leonardo; Passarini, Fabrizio; De Gennaro, Gianluigi; Barbieri, Pierluigi; Licen, Sabina; Perrone, Maria Grazia; Piazzalunga, Andrea; Borelli, Massimo; Palmisani, Jolanda; Di Gilio, Alessia; Rizzo, Emanuele; Colao, Annamaria; Piscitelli, Prisco; Miani, Alessandro

VERSION 1 – REVIEW

REVIEWER	Rima Isaifan Hamad Bin Khalifa University, Qatar.
REVIEW RETURNED	16-Apr-2020

GENERAL COMMENTS	1- An overall comment/limitation is related to the main question on why the authors have used PM10 rather than PM2.5 for the base of their study and hypothesis. It is well documented and proven that PM2.5 is more closely related to respiratory diseases, lung cancer and pre mature mortality due to air pollution. The authors should clearly state this point if it was due to the lack of enough data from the monitoring stations of other reasons. It should be also pointed out in the limitation section. 2- It is critical to give a credit to the first research published that connects the impact of COVID-19 on air quality in the introduction, so kindly cite the following article: Isaifan, R.J., (2020). The dramatic impact of Coronavirus outbreak on air quality: Has it saved as much as it has killed so far. Global Journal of Environmental Science and Management, 6(3): 275-288. https://www.gjesm.net/article_38731.html 3- The methodology section needs major enhancement. Details on how was the doubling time calculated and how was the basic reproductive number estimated should added. This is statistical-based work and the details related the use of each parameter should be clearly stated. 4- Figure 1: (figure 1 a) why the duration of the obtained data was considered for 9-29th of February and is not the same period for the map reporting the spread of COVID-19 infected people in figure 1 (b-e) in March 3-13th. This point is critical as during the second duration where shut down or lockdown has been activated, the PM10 concentration will be much less than the observed level during February and hence, this will greatly impact the results in this work.

	5- More in depth discussion is needed, specially for figure 2 where a clear impact of shutdown is seen by immediate decline in number of infected people the right day after. 6- Where is the discussion of figure 3a? 7- Page 9, line 34: in the value (,19) why did the authors use a comma for the decimal instead of period to be consistent with the rest of numbers in the manuscript? Besides, a comma in French system indicates other values and not decimal at all. 8- Page 9, line 48, the same question above applies to the use of comma in the value 0,27?

REVIEWER	Dario Caro Aarhus University, Denmark
REVIEW RETURNED	23-Apr-2020

GENERAL COMMENTS	The paper is very interesting and crucial for this "special" moment. It is well written and conclusions are important and in line with the results obtained. I really recommend an additional specialist statistical review which i, unfortunately, can't provide. Taking good the statistical analysis, my concern on this paper is: why did the authors only investigate PM10? All introduction of the paper is focused on both PM2.5 and PM10, highlighting more or less recent studies in which particulate matters are connected with the virus. In the title i see "particulate matter" which is general. The background section begins with "Exposures to PM2.5 and PM10 such as those usually recorded in the Po Vally". Suddenly, in the method and results sections, PM2.5 disappears leaving space at PM10 only. So the provocative question could be: did not the author find a correlation with PM2.5 and then they decided to omit them from the analysis? But i don't want to be provocative at all. Because, in my opinion this paper would be convincing also if it shows a non-correlation with PM2.5 and a correlation with PM10. Otherwise, the absence of PM2.5 from the analysis should be properly stated as limitation inviting other researchers to further examine them in depth. Or, explaining why (if there is a reason) you only focused on PM10. Moreover, i invite authors to enlarge their reference list with the additional literature on covid. I understand that this is pretty hard now due to the imminent emerging literature on covid. For this reason i recommend to update the reference list at the very last moment.

REVIEWER	Marie-Rachelle Narcisse, PhD University of Arkansas for Medical Sciences
REVIEW RETURNED	11-May-2020

GENERAL COMMENTS	Title: The Potential role of Particulate Matter in the Spreading of COVID-19 in Northern Italy: First Evidence-based Research Hypotheses Journal: BMJ Open Manuscript ID: bmjopen-2020-039338 Article Type: Original research Authors: Setti, L. et al. Current Study: The authors studied associations between the geographical distribution of daily Particulate Matter (PM10) exceedances and the spread of COVID-19 in the Italian Provinces. While this paper is pertinent and will make an important contribution to the understanding of the spread of COV)D-19, it suffers from considerable pitfalls that, if remedied, will make this manuscript acceptable for publication.

	I kindly make the following recommendations: 1. INTRODUCTION/BACKGROUND: please clearly state the hypotheses. The hypotheses are mentioned in the Discussion section, and in unconventional manner: “Our research hypothesis is that the acceleration of the growth rate observed in Milan could be attributed to a “boost effect” (a kind of exceptional “super-spread event”) on the viral infectivity of COVID-19, corresponding to the peaks of particulate matter.” (p9, lines 57-60) “The other hypothesis is that PM could act as a carrier for droplet nuclei, triggering a boost effect on the spread of the virus” (p. 10, lines 2-4). The hypotheses must arise as a result of previous studies/evidence. The hypotheses must be delineated at the end of the Introduction section before the Methods. This is especially important since your title mentions “First Evidence-based Research Hypotheses”. The hypotheses provide the justification for the choice of the methods used. Furthermore, kindly rephrase the hypotheses. Although you could further explain your hypotheses in a lay term such as “boost effect” of “a kind of exceptional “super-spread event””, the way in which they are stated makes it difficult to justify the analytical plan. 2. INTRODUCTION/BACKGROUND: What are the objectives of this study. This should be clearly stated before the METHODS section. 3. METHODS: It would be edifying to the reader if the authors could describe the study population. They refer to “big cities located in different geographic areas and with remarkable differences in PM10 exceedances, presenting at the same time quite similar urbanization, life style, population, ageing index, and number of commuters” (p. 5, lines 50-55). For readers, not familiar with Italy, it would be difficult to grasp this description fully. Although a detailed portrait of the 110 Provinces is not required, it is crucial to have a better understanding of the population density, urbanization, commuters, ageing index, temperature, pollution, etc., at least for Milan and Rome. In sum, the relevant predictors of the infection rate that should be taken into account when examining the potential role of particulate matter and the spread of the virus should be described. This will help the reader interpret the findings in perspective. 4. METHODS: The only technique mentioned in this section is “recursive binary partitioning tree approach, as implemented into the party package [25] of R [26]” (p. 5, lines 30-31). It would be enlightening to describe this method in light of what the authors wish to accomplish. With regard to performing “statistical inference analyses,” the authors need to be more explicit on the specific statistical methods used and why. It is only then that the reader could discriminate whether the technique used is appropriate as opposed to another one (e.g., time series analysis for example?). Although some references (performing a bivariate conditional regression exploratory analysis, logistic regression, a binomial distributed generalized linear model) are made in the Results section, they are scattered, and should be presented in the METHODS section with clear justification as to why they are used. Why was a semi-logarithmic scale used? How was the time lag handled? It is difficult to justify the choice of the regression analyses delineated in the RESULTS section, as the authors have not made precise the measurement scale of the response variables. 5. METHODS: Please restructure this whole section with sub-sections (e.g., study population, measures, statistical analysis, level of significance).

	6. RESULTS: The authors report: “17,660 infected people during the time-lapse of the study” (p. 5, lines 6-7). Please put this absolute number in perspective. For example, in terms of proportions or per 100,000 inhabitants. 7. RESULTS: The results are not presented in an organized way. It would help if the authors could delimit sub-sections with clear, concise ideas. The results are scattered, and the main hypotheses that the authors would like to present are lost. 8. DISCUSSION: The main findings of the study should be reiterated succinctly at the beginning of the discussion. Furthermore, rather than delineating what has been found in the literature, it is crucial to contrast the results of this study with previous research to appreciate the “plus-value” and innovation of this study. Furthermore, the authors do not provide a robust discussion of the limitations in terms of the variables chosen, potential omitted variables, the analytical strategies. Since COVID-19 is a novel virus, and the authors present some hypotheses, it would be informative that the authors discuss a few potential counter-hypotheses and explain why the ones proposed in this study are more pertinent. I did appreciate the pertinence of this important study. Please resubmit.

REVIEWER	Kefang Lai State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University
REVIEW RETURNED	22-May-2020

GENERAL COMMENTS	It’s my honor to review this interesting and relevant piece of work. Based on the analysis of the number of COVID-19 infected people in 110 Provinces of Italy from February 24th to March 13th (the date when the lockdown has been imposed over Italy) and official daily data relevant to ambient PM10 levels from all Provinces between February 9th and February 29th, the authors found a significant association between the geographical distribution of daily PM10 exceedances and the spread of COVID-19 in the 110 Italian Provinces. This finding is really important to boost environment protection to help stop the spread of SARS-CoV-2. My comments are as follows. Major: Similar finding has been reported by recent studies conducted in China (Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China, PMID: 32315904), America (Correlation Between Environmental Pollution Indicators and COVID-19 Pandemic: A Brief Study in Californian Context, PMID: 32405084), and Italy (Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy, PMID: 32387671). All of the above studies not only focused on the association between PM10 and the spread of COVID-19, but also focused on other pollutants, such as PM2.5, O3, NO, SO2. Why only PM10 was selected in this study? What this study adds to the field? Minor: 1. Kindly shorten the content regarding the association between PM and infectious disease in the Introduction section (page 3 line

	33-page 4 line 27). It's unnecessary to elaborate so many previous studies that were talking about the same thing. 2. Typing errors Page 8 line 34 " 0,19 per day" should be 0.19 per day Page 8 line 48 "0,27 per day" should be 0.27 per day

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Rima Isaifan
Institution and Country:Hamad Bin Khalifa University, Qatar.

1- An overall comment/limitation is related to the main question on why the authors have used PM10 rather than PM2.5 for the base of their study and hypothesis. It is well documented and proven that PM2.5 is more closely related to respiratory diseases, lung cancer and pre mature mortality due to air pollution. The authors should clearly state this point if it was due to the lack of enough data from the monitoring stations of other reasons. It should be also pointed out in the limitation section.

Thank you for this comment that give us the opportunity to improve the manuscript by explaining that and by carrying out also a comparative analysis PM10/PM2.5. Actually, THE Institutional Italian Air Quality MONITORING system is based on widespread network of PM10 monitoring stations and a smaller number of PM2.5 monitoring stations (438 vs 183, respectively). European Directive 2008/50/EU sets a 24 hours threshold for PM10 but not for PM2.5. Therefore, at first instance it has been natural to consider the number of daily threshold exceedances for PM10 concentrations, as foreseen by current law. Stimulated by the precious questions of reviewers and being well aware that high concentrations of PM2.5 show stronger health effects than those of PM10 in epidemiological studies, we proceeded considering the exceedances of available average daily concentrations for PM2.5 with respect of the - not legally compulsory/binding - daily threshold (25 µg/m3 24-hour mean) suggested for PM2.5 by WHO. In this way, we have obtained a confirmation of the significance of the normalized exceedances of daily particulate matter threshold in predicting the infection / population ratio. Frequency of acute exposures appears to be relevant, while no evidence of predictivity was found considering average values or other quartiles in the considered period for PM10 or PM2.5 concentrations. We have explained that in the text of the article
revised. Moreover, we have hypothesized that PM could act as carrier for SARS CoV2, after coagulation of small viral droplets with pre-existing PM; since PM2.5 is in principle physically inclded in PM10, this latter can be considered as an adequate proxy for aerosol viral carrier.

2- It is critical to give a credit to the first research published that connects the impact of COVID-19 on air quality in the introduction, so kindly cite the following article:
Isaifan, R.J., (2020). The dramatic impact of Coronavirus outbreak on air quality: Has it saved as much as it has killed so far. Global Journal of Environmental Science and Management, 6(3): 275-288.
<https://www.gjesm.net/article_38731.html>https://www.gjesm.net/article_38731.html

Thank you for your suggestion. We added this citation in the paper.

3- The methodology section needs major enhancement. Details on how was the doubling time calculated and how was the basic reproductive number estimated should added. This is statistical-based work and the details related the use of each parameter should be clearly stated.

We thank the reviewer and we tried to explain that in this revision. The incidence growth rate in Italy was determined during the first 16 days by the new daily infections (figure 2) taking in account of a typical exponential trend of the spread. The doubling time was calculated with the following usual equation
t1/2 = Ln(2)/(incidence growth rate)
Incidence growth rate is equal to the exponential of the best fitting equation

Now we have introduced a new figure 2b showing the trend of the COVID spread in Italy in the first 16 days, in order to be compared with the specific trend in Milan and Rome presented in Figure 4a.

Figure 2b was cited in the text at page 9 line 1 of the original proof:

"Our study analyzed the Italian outbreak before March 11th when the incidence growth rate showed a typical exponential trend of the spread (Figure 2b)."

Caption of Figure 2 is changed in:

"Figure 2. (a) New daily COVID-19 infections in Italy from February 24th to April 4th, 2020; (b) Trend of spread in Italy in the first 15 days of infection."

About the basic reproductive number (R0), there are not data in literature referred to the single Italian provinces and thus the discussion can be only speculative between the exponential growth rates and doubling time in Italy and China as well as making a comparison with the data in literature on R0 that we have reported (adding a supplemental reference).

4- Figure 1: (figure 1 a) why the duration of the obtained data was considered for 9-29th of February and is not the same period for the map reporting the spread of COVID-19 infected people in figure 1 (b-e) in March 3-13th. This point is critical as during the second duration where shut down or lockdown has been activated, the PM10 concentration will be much less than the observed level during February and hence, this will greatly impact the results in this work.

We tried to better explain that the epidemiological observations of March 3-13 correspond to the contagion occurred in the middle or at the end of February: this is the reason why we had to choose a previous period for the analyses of PM concentration that could have played a role (that is the hypothesis of our work) in enhancing the diffusion of COVID-19.

In Figure 1 we reported the dates in which the exceedances and the infections were really determined; however, we have to considered that the detections of the infected people is affected by the lag period as described at page 6 line 35 of the original proof, for instance: March 3rd of figure 1b corresponds approximately to the incidences of February 15th.

We have changed the caption of the figure 1 in:
"Figure 1. (a) Average daily PM10 exceedances vs. number of monitoring stations in different Italian Provinces from February 9th to 29th, 2020; (b-e) Spreading of COVID-19 infected people detected during the period March 3rd-13th, 2020"

Anyway, PM10 concentration has significantly decreased in the days and weeks following the lockdown, particularly in the Po Valley (see for instance: https://towardsdatascience.com/is-covid-19-lockdown-cleaning-the-skies-over-milan-42dbba1ec812). This is mostly related to the particular configuration of this area, which prevents air circulation and rapid decrease of primary and secondary pollutants.

5- More in depth discussion is needed, specially for figure 2 where a clear impact of shutdown is seen by immediate decline in number of infected people the right day after.

Thanks. To take into account the precious suggestion of the reviewer, we added the following part at page 6 line 1 of the original proof:
"SARS-Cov-2 is a high contagiously virus by airborne direct contact showing super-spread event and then each government measure able to contain the spread should show an immediate reduction of the infection [Seth Flaxman, Swapnil Mishra, Axel Gandy et al. Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London (30-03-2020) doi: https://doi.org/10.25561/77731]. In Figure 2, two main discontinuity trends are evident that can be attributed to the mainItalian government measures to contain the contagion. If continuing the observation beyond the date of the lockdown(March 11th), it was possible……."

6- Where is the discussion of figure 3a?

Thanks. We added the following part at the caption of Figure 3:

"Figure 3. Relationship … (62 out of 66) (c) Boxplots suggest that considering also PM2.5 exceedances rates (despite the 39% presence of missing data) the proportion of COVID-19 in Po Valley might furtherly be stratified ($p < 0.001$) in agreement with previous Figures 1 (b-e)"

7- Page 9, line 34: in the value (19) why did the authors use a comma for the decimal instead of period to be consistent with the rest of numbers in the manuscript? Besides, a comma in French system indicates other values and not decimal at all.
Page 9, line 48, the same question above applies to the use of comma in the value 0,27?

Thanks. We changed the following part in page 6 line 36 of the original proof:

"On the basis, … PM_{10} concentrations: the less polluted Provinces had a median 0.03 infection case over 1000 residents (first – third quartile 0.01 – 0.09, range 0.00 – 0.56), while most polluted Provinces had a median 0.26 infection cases over 1000 Province residents (first – third quartile 0.14 – 0.51, range 0.00 – 4.92). The boxplots in Fig. 3b are log-transformed to enhance figure readability."

We added the following part in page 7 line 61 of the original proof:

"PM2.5 exceedances have been also explored: correlation between Province PM2.5 and PM10 exceedance rates has been assessed by a linear model (intercept 1.06, slope 1.38), with satisfactory accuracy ($p < 0.001$, mild lack of normality into residuals according diagnostic plots, residual standard error: 1.82). PM2.5 exceedance rates appear to be very highly correlated to PM10 exceedance rates: Pearson coefficient 0.94 (95% percent confidence interval: 0.90 – 0.96).
Repeating exploratory analysis on PM2.5 exceedance rates by recursive partitioning tree analysis, an 11 cutoff was detected and an improved recursive partitioning tree was disclosed, as depicted in Figure 3C. Despite the limitation of not having 43 over 110 complete data on PM2.5 exceedances (39% missing values), it appears that, coherently with main study outcome, strong PM2.5 exceedance rates are eminently located in Po Valley."

The following table depicts in detail the situation:

exceedance rates	north	south

PM_{10}<= 1.29	2	28
PM_{10}>1.29 , $PM_{2.5}$<= 11	18	2
$PM_{2.5}$> 11	16	0

Reviewer: 2
Reviewer Name:Dario Caro
Institution and Country:Aarhus University, Denmark

1. The paper is very interesting and crucial for this "special" moment. It is well written and conclusions are important and in line with the results obtained. I really recommend an additional specialist statistical review which, unfortunately, can't provide. Taking good the statistical analysis, my concern on this paper is: why did the authors only investigate PM10? All introduction of the paper is focused on both PM2.5 and PM10, highlighting more or less recent studies in which particulate matters are connected with the virus. In the title i see "particulate matter" which is general. The background section begins with "Exposures to PM2.5 and PM10 such as those usually recorded in the Po Vally". Suddenly, in the method and results sections, PM2.5 disappears leaving space at PM10 only.

So the provocative question could be: did not the author find a correlation with PM2.5 and then they decided to omit them from the analysis? But i don't want to be provocative at all.
Because, in my opinion this paper would be convincing also if it shows a non-correlation with PM2.5 and a correlation with PM10. Otherwise, the absence of PM2.5 from the analysis should be properly stated as limitation inviting other researchers to further examine them in depth. Or, explaining why (if there is a reason) you only focused on PM10.

Thank you for this comment that give us the opportunity to refer also to PM 2.5 thus improving the quality of the paper. We really thank the reviewer.
Institutional Italian Air Quality system deploys a widespread network of PM10 monitoring stations and a smaller number of PM2.5 monitoring stations (438 vs 183). European Directive 2008/50/EU sets a 24 hours threshold for PM10 but not for PM2.5, so at first instance it has been natural to consider the number of daily threshold exceedances for PM10 concentrations, as foreseen by current law in force. Stimulated by the precious questions of reviewers and being well aware that high concentrations of PM2.5 show stronger health effects than those of PM10 in epidemiological studies, we proceeded considering the exceedances of available average daily concentrations for PM2.5 with respect of the - not legally binding - daily threshold (25 µg/m3 24-hour mean) suggested for PM2.5 by WHO. We obtained a confirmation of the significance of the normalized exceedances of daily particulate matter threshold in predicting the infection / population ratio. Frequency of acute exposures appears to count while no evidence of predictivity was found considering averages or other quantiles in the period for PM10 or PM2.5 concentrations.
Moreover we hypothesize that PM could act as carrier for SARS CoV2, after coagulation of small viral droplets with pre-existing PM; since PM2.5 is in principle physically included in PM10, PM10 can be considered as an adequate proxy for aerosol viral carrier.

2. Moreover, I invite authors to enlarge their reference list with the additional literature on covid. I understand that this is pretty hard now due to the imminent emerging literature on covid. For this reason i recommend to update the reference list at the very last moment.

We added the following part in page 3 line 46 of the original proof:

“Recently, we have evidenced the first world evidence of the presence of COVID-19 on outdoor PM in samples tested in the province of Bergamo (Lombardy, Italy), which experienced the highest diffusion and mortality rates in Italy, in samples harvested between February 23th and March 9th [Setti et al (2020) Environmental research].”

We added the following part in page 3 line 58 of the original proof:

“This article presents the data that led to the publication of the Position Paper and triggered high interest of the research community at working on this topic suggesting a further trasmission via airborne dust [Guangbo Qu, Xiangdong Li, Ligang Hu, and Guibin Jiang , An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19), Environmental Science & Technology 2020 54 (7), 3730-3732, DOI: 10.1021/acs.est.0c01102] even if the survival of the virus depends on the environmental conditions such as humidity and temperature [Sharma, V.K., Jinadatha, C. &Lichtfouse, E. Environmental chemistry is most relevant to study coronavirus pandemics. Environ ChemLett (2020). https://doi.org/10.1007/s10311-020-01017-6]. Other papers support the possible merging of contaminated aerosol with fine particulate in the atmosphere [J. K. Mutuku, W.-C. Hou, W.-H. Chen, “An Overview of Experiments and Numerical Simulations on Airflow and Aerosols Deposition in Human Airways and the Role of Bioaerosol Motion in COVID-19 Transmission”, Aerosol and Air Quality Research, 20: 1172–1196, 2020; J. Borak, “Airborne Transmission of COVID-19”, Occupational Medicine, https://doi.org/10.1093/occmed/kqaa080; M. A. Zoran, R. S. Savastru, D. M. Savastru, M. N. Tautan, “Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy”, Science of The Total Environment, 738, 139825, 2020].
The concentration of fine particles has been also repeatedly recognized by other authors as an important co-factor in the mortality level in highly contaminated areas [E. Conticini, B. Frediani, D. Caro, “Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy?”, Environmental Pollution, 261, 114465, 2020; R. Pansini, D. Fornacca, “Initial evidence of higher morbidity and mortality due to SARSCoV-2 in regions with lower air quality”. medRxivhttps://doi.org/10.1101/2020.04.04.20053595, 2020; X. Wu, R.C. Nethery, B. Sabath, D. Craun, F. Dominici, “Exposure to air pollution and COVID-19 mortality in the United States”, medRxiv, doi:10.1101/2020.04.05.20054502, 2020]”

We also added the following references:

26. Domingo JL, Rovira J. Effects of air pollutants on the transmission and severity of respiratory viral infections. EnvironmentalResearch. 2020 May 11:109650.
27. Sanità di Toppi L, Sanità di Toppi L, Bellini E. Novel coronavirus: How atmospheric particulate affects our environment and health. Challenges. 2020 Jun;11(1):6.
Ref. 42 Distante C, Piscitelli P, Miani A. Covid-19 outbreak progression in Italian regions: Approaching the peak by the end of March in northern Italy and first week of April in Southern Italy. International Journal of Environmental Research and Public Health. 2020 Jan;17(9):3025.
Reviewer: 3
Reviewer Name: Marie-Rachelle Narcisse, PhD
Institution and Country@ University of Arkansas for Medical Sciences

1. INTRODUCTION/BACKGROUND: please clearly state the hypotheses. The hypotheses are mentioned in the Discussion section, and in unconventional manner: “Our research hypothesis is that the acceleration of the growth rate observed in Milan could be attributed to a “boost effect” (a kind of exceptional “super-spread event”) on the viral infectivity of COVID-19, corresponding to the peaks of particulate matter.” (p9, lines 57-60) “The other hypothesis

is that PM could act as a carrier for droplet nuclei, triggering a boost effect on the spread of the virus" (p. 10, lines 2-4). The hypotheses must arise as a result of previous studies/evidence. The hypotheses must be delineated at the end of the Introduction section before the Methods. This is especially important since your title mentions "First Evidence-based Research Hypotheses". The hypotheses provide the justification for the choice of the methods used. Furthermore, kindly rephrase the hypotheses. Although you could further explain your hypotheses in a lay term such as "boost effect" of a kind of exceptional "super-spread event", the way in which they are stated makes it difficult to justify the analytical plan.

This has been defined an evidence-based research hypotheses because we have developed a model to suggest a super-spread event as a consequence of the experimental data discussion. However, the editor asked us to change the title in order to include information about the study type and therefore and we had to replace "First Evidence based hypotheses" with "First Observational Study"

We have inserted the requested sentence at the end of the introduction : The research hypotheses that we addressed is the possibility that air pollution could produce a "boost effect" of COVID 19 epidemic, thus resulting a kind of exceptional "super-spread event".

In order to properly address the points highlighted by the reviewer, we changed the following part in page 10 line 25 to 56 of the original proof:

"Our findings suggested that the acceleration of the growth rate observed in Milan could be attributed to a "boost effect" (a kind of exceptional "super-spread event") on the viral infectivity of COVID-19, corresponding to the peaks of particulate matter. According to this hypothesis, PM could then act as a carrier for droplet nuclei, … and promoting contagion. These first observations suggest that particulate matter could be regarded as ancontributing factor to the severity of COVID-19 infection in terms of airborne diffusion and health outcomes."

2. INTRODUCTION/BACKGROUND: What are the objectives of this study. This should be clearly stated before the METHODS section.

We thank the reviewer and we expanded the introduction, also according to the suggestion of another reviewer, adding other publications in the reference list.

Furthermore, we added the following part in page 3 line 58 of the original proof:

"This article presents the data that led to the publication of the Position Paper and triggered high interest of the research community at working on this topic suggesting a further trasmission via airborne dust [Guangbo Qu, Xiangdong Li, Ligang Hu, and Guibin Jiang , An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19), Environmental Science & Technology 2020 54 (7), 3730-3732, DOI: 10.1021/acs.est.0c01102] even if the survival of the virus depends on the environmental conditions such as humidity and temperature [Sharma, V.K., Jinadatha, C. &Lichtfouse, E. Environmental chemistry is most relevant to study coronavirus pandemics. Environ Chem Lett (2020). https://doi.org/10.1007/s10311-020-01017-6]. The focus of this study is to present a thorough statisticalanalysis between the COVID-19 infectionsinItaly at the end of February-first of March 2020 (February 24th-March 13th) and the frequency ofhighdaily average concentrations of PM occurring the days before (February 9th-29th) taking into account the lag period.By analyzing the relationship found between COVID-19 outbreaks and PM levels on a Province base, the aim is that to interpret the super-spread of COVID-

19 outbreaks registered in the North Italy and reveal the possible contribution of PM levels in the virus airborne transmission."

3. METHODS: It would be edifying to the reader if the authors could describe the study population. They refer to "big cities located in different geographic areas and with remarkable differences in PM10 exceedances, presenting at the same time quite similar urbanization, life style, population, ageing index, and number of commuters" (p. 5, lines 50-55). For readers, not familiar with Italy, it would be difficult to grasp this description fully. Although a detailed portrait of the 110 Provinces is not required, it is crucial to have a better understanding of the population density, urbanization, commuters, ageing index, temperature, pollution, etc., at least for Milan and Rome. In sum, the relevant predictors of the infection rate that should be taken into account when examining the potential role of particulate matter and the spread of the virus should be described. This will help the reader interpret the findings in perspective.

According to the precious suggestions of the reviewer, we changed the following part in page 4 line 16 to 47 of the original proof:

"Population data related to each Italian Province were collected from ISTAT database for the 110 Provinces [3]: the population absolute frequencies and their density within Provinces (n° inhabitants/km^2); the number of commuters and its proportion with respect to the overall Province population. The primary outcome of our study is the number of COVID-19 infected people for each Province from February 24th to March 13th (the date when the lockdown was decided), as reported on the official Government website, updated with daily frequency [24]: the main response variable is therefore the Province infection rate of the disease, expressed as a proportion. PM exceedances were collected between February 9th and February 29th, taking intoccount the lag period, which is the average time elapsed between the initial infection and the diagnosis. Further covariates related to the different Provinces have also been collected into our study dataset: the number of the air quality monitoring stations present in each Province, the longitude and the latitude of the Province cities center, the two levels (north, south) geographical Province classification.
The dataset is publicly available at http://www.biostatisticaumg.it/bmj/; at the same web page it is also available the statistical analysis reproducible code written in R language [26].
To investigate how PMexceedances might relate to infection diffusion, we started performing an exploratory analysis on PM_{10} exceedances considering the recursive partitioning tree approach, as implemented into the party package [25]. Such implementation connects the exploratory techniques to the classical statistical test approach, with the advantage to exploit a motivated stopping criterion when pruning the tree – i.e. the p-value of a significance test on independence on any of the covariate and the response [Everitt, B. S., Hothorn, T. A Handbook of Statistical Analyses using R. Chapman & Hall, CRC Press (2006)]. Within recursive partitioning analysis the response variable was the proportion of COVID-19 cases over Province population; to improve graphical readability, Figures depict the log-transform of such proportions response. Cut-offs identified by the recursive partitioning tree analysis where subsequently used into the – both univariate and multivariable - binomial generalized linear models, i.e. the logistic regression. The response of the binomial generalized linear models is expressed as a two-dimensional vector [Crawley, M. J. Statistics: An Introduction Using R, 2nd Edition. Wiley (2014)] obtained binding the number of COVID-19 cases and the rest of the Province population. In presence of overdispersion, quasibinomial distributions were addressed [40massimo]. When suitable, association in contingency table has been expressed also in terms of odds ratios, and the Fisher exact test was issued to asses statistical significance.
Correlation between Province $PM_{2.5}$ and PM_{10} exceedances rates has been addressed by a linear model. Pearson coefficient was used to evaluate correlation and diagnostics plots were issued to assess model adequacy. Exploratory analysis on $PM_{2.5}$ exceedances rates was held by recursive partitioning tree approach too. We have also performed statistical … of commuters."

We added the following part in page 4 line 56 of the original proof:

“The municipality of Rome is far more extensive, with 1,287 square kilometers of surface compared to just 182 in Milan.If we talk about population, Rome has 2.87 million inhabitants compared to 1.35 in Milan, but it is much less densely populated: in every square kilometer in Milan we find 7,439 inhabitants, while in Rome 2,232, again due to the its extension. However, 0.91 million inhabitants in Rome live in neighborhoods with an average density between 6720 and 9231 inhab. / Km2. The development of the metropolitan networks of Milan and Rome led to a structure today of 98 and 54 km respectively. Rome is confirmed as the main destination with around 29 million presences with respect to Milan, which account for over half of tourists with 12.1 million. The number of daily commuters are higher in Rome (2,04 million trips) compared to Milan (1.66 million trips).”

4. METHODS: The only technique mentioned in this section is “recursive binary partitioning tree approach, as implemented into the party package [25] of R [26]” (p. 5, lines 30-31). It would be enlightening to describe this method in light of what the authors wish to accomplish. With regard to performing “statistical inference analyses”, authors need to be more explicit on the specific statistical methods used and why. It is only then that the reader could discriminate whether the technique used is appropriate as opposed to another one (e.g., time series analysis for example?). Although some references (performing a bivariate conditional regression exploratory analysis, logistic regression, a binomial distributed generalized linear model) are made in the Results section, they are scattered, and should be presented in the METHODS section with clear justification as to why they are used. Why was a semi-logarithmic scale used? How was the time lag handled? It is difficult to justify the choice of the regression analyses delineated in the RESULTS section, as the authors have not made precise the measurement scale of the response variables.
5.

Authors thanks the reviewer for these proper observations. The
whole methods section has been revised and extensively explicited. In particular, we added a couple of textbooks as a reference to provide a theoretical framework to the technique adopted. We choose to explore the dataset by means of regression tree because such a technique is, in a sense, not dependent from researcher's choice of predictor added into a model. Of course, we agree that PM and Covid cases might be modelled, longitudinally, as time series in a very natural way.
But focusing only to an early phase of the disease spread, PM has been summarized into a 'sufficient statistic' (i.e. the exceedances ratios for each Province), reducing the 'longitudinal' scheme to a 'cross-section' view.
In the methods section we have also better explained that the semi-logarithmic scale has been used to provide more readable boxplots; nevertheless, the analysis can be
equivalently performed without any log transform, yielding to the same conclusions – the code and the data are now available at the addresshttp://www.biostatisticaumg.it/bmj/ and the
response has been now better clarified.

6. METHODS: Please restructure this whole section with sub-sections (e.g., study population, measures, statistical analysis, level of significance).

We have restructured the section with the following sub-sections:
Data about daily PM10 exceedances and COVID-19 confirmed infections
Statistical analyses
Sites description for the statistical analyses about Milan and Rome

7. RESULTS: The authors report: “17,660 infected people during the time-lapse of the study” (p. 5, lines 6-7). Please put this absolute number in perspective. For example, in terms of proportions or per 100,000 inhabitants.

We added the following part in page 5 line 17 of the original proof:

“Overall, there were 17,660 infected people on 60,4 million inhabitants in Italy during the time lapse of the study.”

8. RESULTS: The results are not presented in an organized way. It would help if the authors could delimit sub-sections with clear, concise ideas. The results are scattered, and the main hypotheses that the authors would like to present are lost

Results have been slightly extended and 2 subsections are proposed, namely:

“COVID-19 diffusion in Italy and Particulate Matter exceedances” - where we present the results of COVID 19 outbreak at the beginning of march taking into account the lag period of 14-17 days between infection event and diagnosis. Univariate models, bivariate conditional regression exploratory analysis, and final binomial distributed generalized linear model, corrected for overdispersion are presented. PM_{10} exceedances are highly significant predictors for registered number of infections over 1000 inhabitants for each Italian province. Number of commuters and PM2.5 exceedances of WHO daily guidelines are discussed as well.

“Understanding and Modeling the Super-spreading Events in Milan and Rome” - The contrast between the infection spreads in two Italian megalopolis is proposed for supporting the relevance of airborne route for SARS CoV2 in Milan but not in Rome.

9. DISCUSSION: The main findings of the study should be reiterated succinctly at the beginning of the discussion. Furthermore, rather than delineating what has been found in the literature, it is crucial to contrast the results of this study with previous research to appreciate the “plus-value” and innovation of this study. Furthermore, the authors do not provide a robust discussion of the limitations in terms of the variables chosen, potential omitted variables, the analytical strategies. Since COVID-19 is a novel virus, and the authors present some hypotheses, it would be informative that the authors discuss a few potential counter-hypotheses and explain why the ones proposed in this study are more pertinent.

By Reviewing all the paper taking in consideration the suggestions of the referee, we have better explained the objectives of our findings in the introduction as reported in the answer N. 2

Reviewer: 4
Reviewer Name: Kefang Lai
Institution and Country: State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University

1. Similar finding has been reported by recent studies conducted in China (Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China, PMID: 32315904), America (Correlation Between Environmental Pollution Indicators and COVID-19 Pandemic: A Brief Study in Californian Context, PMID: 32405084), and Italy (Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy, PMID: 32387671). All of the

above studies not only focused on the association between PM10 and the spread of COVID-19, but also focused on other pollutants, such as PM2.5, O3, NO, SO2. Why only PM10 was selected in this study? What this study adds to the field?

Institutional Italian Air Quality system deploys a widespread network of PM10 monitoring stations and a smaller number of PM2.5 monitoring stations (438 vs 183). European Directive 2008/50/EU sets a 24 hours threshold for PM10 but not for PM2.5, so at first instance it has been natural to consider the number of daily threshold exceedances for PM10 concentrations, as foreseen by current law in force. Stimulated by the precious questions of reviewers and being well aware that high concentrations of PM2.5 show stronger health effects than those of PM10 in epidemiological studies, we proceeded considering the exceedances of available average daily concentrations for PM2.5 with respect of the - not legally binding - daily threshold (25 µg/m3 24-hour mean) suggested for PM2.5 by WHO. We obtained a confirmation of the significance of the normalized exceedances of daily particulate matter threshold in predicting the infection / population ratio. Frequency of acute exposures appears to count while no evidence of predictivity was found considering averages or other quantiles in the period for PM10 or PM2.5 concentrations.
Moreover we hypothesize that PM could act as carrier for SARS CoV2, after coagulation of small viral droplets with pre-existing PM; since PM2.5 is in principle physically included in PM10, PM10 can be considered as an adequate proxy for aerosol viral carrier.The coagulated viral particles reduce their diffusion coefficient and persist more in the atmosphere. The other pollutants are unable to play this role

2. Kindly shorten the content regarding the association between PM and infectious disease in the Introduction section (page 3 line 33-page 4 line 27). It is unnecessary to elaborate so many previous studies that were talking about the same thing.

We have re-written and revised the section

3. Typing errors

Page 8 line 34 “0,19 per day” should be 0.19 per day
Page 8 line 48 “0,27 per day” should be 0.27 per day

We have re-written and revised the section

VERSION 2 – REVIEW

REVIEWER	Rima Isaifan HBKU
REVIEW RETURNED	28-Jun-2020

GENERAL COMMENTS	The authors have responded to the comments efficiently.

REVIEWER	Dario Caro Aarhus University (Denmark)
REVIEW RETURNED	29-Jun-2020

GENERAL COMMENTS	Authors prompty replied to all my comments. I believe that the manuscript has been sufficiently improved. From my side, the paper is ready to be published.

REVIEWER	Marie-Rachelle Narcisse, PhD University of Arkansas for Medical Sciences, Fayetteville. United States of America

REVIEW RETURNED	20-Jul-2020

GENERAL COMMENTS	The authors have done a great job revising the manuscript and addressing reviewers' concerns. Your contribution is crucial and edifying. Marie-Rachelle Narcisse, PhD

REVIEWER	Kefang Lai State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University
REVIEW RETURNED	13-Jul-2020

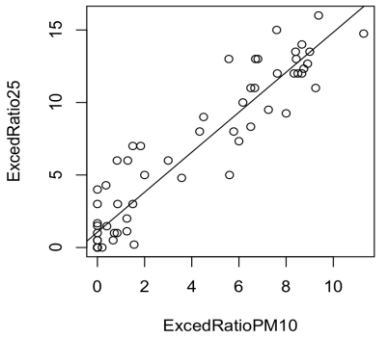
GENERAL COMMENTS	The revised manuscript improved a lot compared to its original version, however, a few major questions have not been addressed yet. 1. Although the association between PM2.5 and PM10 exceedances was analyzed, the contribution of PM2.5 to COVID-19 infection was not mentioned. In addition, only univariate analysis was applied, why bivariate analysis (PM2.5 combined with PM10) was not used? Will the result change if the PM10 daily limit value is adjusted by PM2.5 daily limit value? Despite the fact that PM2.5 data was not available in 43 provinces out of 110, I suggest that bivariate analysis could be considered in provinces who have full data, otherwise, it should be pointed out as a limitation in the discussion section. 2. In the methodology section, details on how the doubling time and the basic reproductive number were calculated have not been added. 3. There are many spelling mistakes in the manuscript. For example, Page 3, line 33: COVID19 should change to COVID-19; Page 5, line 61: PM10monitoring should be changed into PM10 monitoring; ...

VERSION 2 – AUTHOR RESPONSE

Reviewer: 4

1. Although the association between PM2.5 and PM10 exceedances was analyzed, the contribution of PM2.5 to COVID-19 infection was not mentioned. In addition, only univariate analysis was applied, why bivariate analysis (PM2.5 combined with PM10) was not used? Will the result change if the PM10 daily limit value is adjusted by PM2.5 daily limit value? Despite the fact that PM2.5 data was not available in 43 provinces out of 110, I suggest that bivariate analysis could be considered in provinces who have full data, otherwise, it should be pointed out as a limitation in the discussion section.

We thank the reviewer for the observation. During the first revision, according to the R code published (http://www.biostatisticaumg.it/bmj/codeReproducible.txt) we investigate in a bivariate way the association between PM2.5 and PM10 exceedances, observing a strong rho = 0.94 (95 percent conf int. 0.90 - 0.96, t = 22.1, p < 0.001)

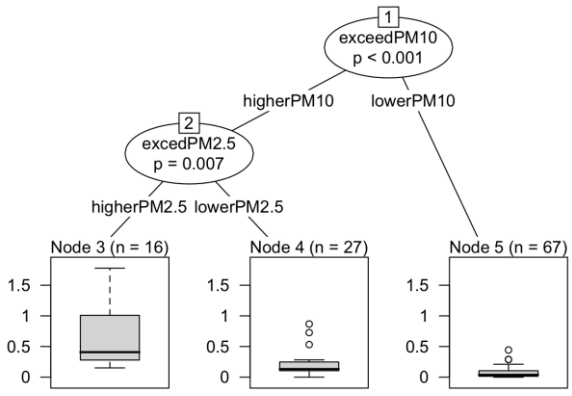


and we confirm that during the first revision, according to the R code published (http://www.biostatisticaumg.it/bmj/codeReproducible.txt) we did use bivariate analysis (PM2.5 combined with PM10) in the Relation3C. The line codes are:

```
### Repeating exploratory analysis on PM2.5 exceedance rates by
### recursive partitioning tree analysis, an 11 cutoff was detected ...
monovariatePM25 = LogCasesProp ~  ExcedRatio25
plot(ctree(monovariatePM25))
excedPM2.5 = factor(ExcedRatio25 > 11)
levels(excedPM2.5)[2] = "higherPM2.5"
levels(excedPM2.5)[1] = "lowerPM2.5"

### ...  and an improved tree was disclosed, as depicted in Figure 3C.
Relation3C = LogCasesProp ~ exceedPM10 + excedPM2.5 + CommutersDensity
Figure3C = ctree(Relation3C)
plot(Figure3C)
```

and we obtained the so called Figure 3C:



We see that the Figure 3C discloses the main role of lower PM10 excedancees levels (according to the cut-off factor ExcedRatioPM10 > 1.285714) in 67 cities, and only in the higher PM10 cities the PM2.5 becomes significant, according to a ExcedRatio25 > 11 cut-off.

As a consequence, despite the limitation of not having 43 out of 110 complete data on PM2.5 exceedances (due to the absence of PM2.5 monitoring stations in 39% of Italian provinces) we considered the following code:

```
summary(excedPM2.5)
length(excedPM2.5)
```

```
summary(excedPM2.5)/length(excedPM2.5)
threelevelsclass = factor(as.numeric(exceedPM10) * as.numeric(excedPM2.5))
levels(threelevelsclass)[1] = "lowbothPM"
levels(threelevelsclass)[2] = "highPM10lowPM2.5"
levels(threelevelsclass)[3] = "highbothPM"
table(threelevelsclass, Where)
```

in which it appears that, coherently with main study outcome, strong PM2.5 exceedance rates are eminently located in Po Valley, as reported in the table(threelevelsclass, Where):

```
                Where
threelevelsclass   north south
  lowbothPM            2    28
  highPM10lowPM2.5    18     2
  highbothPM          16     0
```

We confirm to the reviewer that the results of our analyses could change if the PM10 daily limit value was adjusted by PM2.5 daily limit value. We recall that those cutoffs are not choosen by authors; they are selected from the recursive algorithm itself in order to minimize variances in a hierarchical manner. Therefore, a classical analysis by a generalized linear model could yield to a less accurate result, but anyway coherent with our main results:

```
suggestedattempt = glm(response ~ ExcedRatioPM10 + ExcedRatio25 + CommutersDensity, family =
quasibinomial)
summary(suggestedattempt)
```

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -13.37773    3.93859  -3.397   0.0012 **
ExcedRatioPM10   -0.04008    0.12668  -0.316   0.7528
ExcedRatio25      0.15989    0.09040   1.769   0.0819 .
CommutersDensity  0.08360    0.08045   1.039   0.3028
```

The model, in a sense, confirms that higher levels of PM2.5 are associated to the response, but the opposite signes (-0.04 versus 0.16) in effect size of PM10 and PM2.5 are not interpretable in a natural way. The simple explanation may rely on the fact that, in a sense, a generalized linear model is 'greedy' and tries to explain the response variability in the straigthforward manner, not in a hierarchical approach. Therefore authors are convinced that such a model should not be considered into the manuscript.

2. In the methodology section, details on how the doubling time and the basic reproductive number were calculated have not been added.

We thank the reviewer and we tried to explain that in this revision. The incidence growth rate in Italy was determined during the first 16 days by the new daily infections (figure 2) taking in account of a typical exponential trend of the spread. The doubling time was calculated with the following usual equation

$t_{1/2} = Ln(2)/(\text{incidence growth rate})$

Incidence growth rate is equal to the exponential of the best fitting equation

Now we have introduced a new figure 2b showing the trend of the COVID spread in Italy in the first 16 days, in order to be compared with the specific trend in Milan and Rome presented in Figure 4a.

Figure 2b was cited in the text at page 9 line 1 of the original proof:

“Our study analyzed the Italian outbreak before March 11th when the incidence growth rate showed a typical exponential trend of the spread (Figure 2b).”

Caption of Figure 2 is changed in:

“Figure 2. (a) New daily COVID-19 infections in Italy from February 24th to April 4th, 2020; (b) Trend of spread in Italy in the first 15 days of infection.”

About the basic reproductive number (R0), there are not data in literature referred to the single Italian provinces and thus the discussion can be only speculative between the exponential growth rates and doubling time in Italy and China as well as making a comparison with the data in literature on R0 that we have reported (adding a supplemental reference).

3. There are many spelling mistakes in the manuscript. For example, Page 3, line 33: COVID19 should change to COVID-19; Page 5, line 61: PM10monitoring should be changed into PM10 monitoring; ...

Answer: unfortunately there should be something happening in the converision of the Word file when re-opening from other computers. We have provided a revision of the spelling of the article. Many thanks.